

Dynamic Iterative Pursuit

Dave Zachariah, Saikat Chatterjee and Magnus Jansson

Abstract—For compressive sensing of dynamic sparse signals, we develop an iterative pursuit algorithm. A dynamic sparse signal process is characterized by varying sparsity patterns over time/space. For such signals, the developed algorithm is able to incorporate sequential predictions, thereby providing better compressive sensing recovery performance, but not at the cost of high complexity. Through experimental evaluations, we observe that the new algorithm exhibits a graceful degradation at deteriorating signal conditions while capable of yielding substantial performance gains as conditions improve.

I. INTRODUCTION

Compressive Sensing (CS) [1] problems assume a sparse-signal model, undersampled by a linear measurement process. The algorithms for CS can be separated into three broad classes: convex relaxation, Bayesian inference, and iterative pursuit (IP). For large-dimensional CS signal-reconstruction, IP algorithms offer computationally efficient solutions. Examples of such IP algorithms are orthogonal matching pursuit (OMP) [2], subspace pursuit (SP) [3] and several variants of them [4][5][6][7]. The methodology of such IP algorithms is to detect and reconstruct the non-zero, or ‘active’, signal coefficients in a least-squares framework. These algorithms may use some prior information, such as the maximum allowable cardinality of the ‘support set’. The support set is defined as the set of active signal coordinates of the underlying sparse signal. In general, the IP algorithms work with a single snapshot of the measurements. In this paper, we are interested in generalizing the iterative pursuit approach so as to use more prior information. Such prior information may for instance be available in dynamically evolving sparse processes with temporal/spectral/spatial correlations, as in the sparse signal scenarios of magnetic resonance imaging (MRI) [8], [9], spectrum sensing [10] and direction of arrival estimation [11].

Incorporation of prior information is a recent trend in CS. In [12], the overall methodology is sequential and can be seen as a two-step approach: (1) support-set detection of the sparse signal, and (2) reduced-order recovery using prior information on the detected support set. For a reasonable detection of support set, [12] uses convex relaxation algorithms. Then, a standard Kalman filter (KF) is employed to use prior information for sequential signal recovery. Without explicit support set detection, [13] uses KF to estimate the entire signal and enforces sparsity by imposing an approximate norm constraint. However, the work of [13] validates their algorithm for a signal with a static sparsity pattern (i.e. an unknown pattern that does not evolve over time). Similarly, [14] considers scenarios with static sparsity patterns and solves the reconstruction of a temporally evolving sparse signal with multiple measurement vectors in a batch Bayesian learning framework with unknown model parameters. Iterative pursuit algorithms that can use prior information to recover dynamic sparse signals are, however, largely unexplored. One exception is [15] which uses a maximum a posteriori criterion to modify SP for Gaussian processes. Their signal model, however, does not allow explicit modeling of the temporal correlation of the sparsity pattern.

In this paper, we consider a signal model with dynamically evolving sparsity pattern. In other words, we consider that the signal sparsity pattern varies over time/space at any rate (i.e. from a

slowly varying case to a rapidly varying case). We then develop a predictive orthogonal matching pursuit algorithm that can incorporate prior information in a stochastic framework, using the signal to prediction error ratio as a statistic. Thereby recovery performance can be improved while maintaining the complexity advantage of IP algorithms. This generalizes the linear minimum mean square error (MMSE) approach taken in Gaussian-based matching pursuit [16].

We also develop a robust detection strategy for finding the support set elements of a dynamic sparse signal. Compared to standard correlation-based successive detection in existing iterative pursuit algorithms, such as OMP, this detection strategy is found to be more robust to erratic changes in the sparsity pattern.

Finally, the algorithm is integrated into a recursive Kalman-filter framework in which the sparse process is predicted as a superposition of state transitions. The new IP algorithm using sequential predictions is referred to as dynamic iterative pursuit (DIP). Through experimental simulations we show that the new algorithm provides a graceful degradation at higher measurement noise levels and/or lower measurement signal dimensions, while capable of yielding substantial gains at more favorable signal conditions.

Notation: $\|\mathbf{x}\|_0$ denotes l_0 ‘norm’, i.e. the number of non-zero coefficients of the vector \mathbf{x} . $\mathbf{A} \oplus \mathbf{B}$ is the direct sum of matrices. $|S|$ and S^c are the cardinality and complement of set S , respectively. \emptyset denotes the empty set. $(\cdot)^*$ is the Hermitian transpose operator. \mathbf{A}^\dagger the Moore-Penrose pseudoinverse of matrix \mathbf{A} . $\mathbf{C}^{1/2}$ denotes a matrix square root of a positive definite matrix \mathbf{C} , and $\mathbf{C}^{*/2}$ is its Hermitian transpose. $\mathbf{A}_{[\mathcal{I}, \mathcal{J}]}$ denotes a submatrix of \mathbf{A} with elements from row and column indices listed in ordered sets \mathcal{I} and \mathcal{J} . Similarly, the column vector $\mathbf{x}_{[\mathcal{I}]}$ contains the elements of \mathbf{x} with indices from set \mathcal{I} .

II. SIGNAL MODEL

We consider a standard CS measurement setup,

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{n}_t \in \mathbb{C}^M, \quad (1)$$

where $\mathbf{x}_t \in \mathbb{C}^N$ is the sparse state vector to be estimated and the \mathbf{n}_t is zero-mean Gaussian, $\mathbb{E}[\mathbf{n}_t \mathbf{n}_{t-l}^*] = \mathbf{R}_t \delta(l)$. The sensing matrix $\mathbf{H} = [\mathbf{h}_1 \ \dots \ \mathbf{h}_N] \in \mathbb{C}^{M \times N}$, where $M < N$. Both \mathbf{H} and \mathbf{R}_t are given. Without loss of generality, we assume $\|\mathbf{h}_i\|_2 = 1$.

A. Process model

Let the ‘support set’ $I_{x,t} \subset \{1, \dots, N\}$ represent the sparsity pattern of $\mathbf{x}_t \in \mathbb{C}^N$. It will be assumed that $|I_{x,t}| \equiv \|\mathbf{x}_t\|_0 \leq K_{\max}$, where $K_{\max} < M$. Let λ_{ji} denote the state transition probability $j \rightarrow i$ of the ‘active’ signal coordinate j . Then the probabilities determine the transition $I_{x,t} \rightarrow I_{x,t+1}$, as will be illustrated below.

The transition of an active signal coordinate $j \rightarrow i$ is modeled as an autoregressive (AR) process,

$$x_{i,t+1} = \alpha_{ij} x_{j,t} + w_{i,t}, \quad (2)$$

where $x_{j,t}$ denotes the j th component of \mathbf{x}_t , $w_{i,t}$ is the associated innovation and the AR coefficient $|\alpha_{ij}| < 1$. This model extends the scenario considered in [14] where the transition probabilities are degenerate $\lambda_{ji} = \delta(i-j)$, resulting in a static sparsity pattern $I_{x,t} \equiv I_x, \forall t$.

The sparse-signal process can be written compactly as a linear state-space model with random transition matrices \mathbf{A}_t and \mathbf{B}_t ,

$$\mathbf{x}_{t+1} = \mathbf{A}_t \mathbf{x}_t + \mathbf{B}_t \mathbf{w}_t, \quad (3)$$

where \mathbf{w}_t is zero-mean Gaussian, $\mathbb{E}[\mathbf{w}_t \mathbf{w}_{t-l}^*] = \mathbf{Q} \delta(l) \in \mathbb{C}^{N \times N}$ and $\mathbf{Q} = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$. The non-zero elements of $\mathbf{A}_t \in \mathbb{C}^{N \times N}$

The authors are with the ACCESS Linnaeus Centre, KTH-Royal Institute of Technology, Stockholm. E-mail: {dave.zachariah,magnus.jansson}@ee.kth.se and saikatchatt@gmail.com. This work was partially supported by the Swedish Research Council under contract 621-2011-5847.

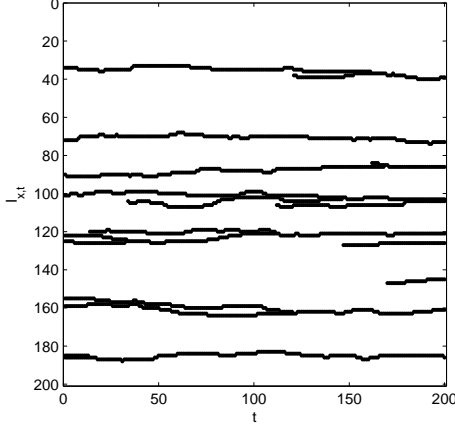


Fig. 1. Example of evolving sparsity pattern with $(N, K, T) = (200, 10, 200)$ and transition probabilities (4).

are $a_{ij,t} = \alpha_{ij}$ for all $j \in I_{x,t}$ and $i \in I_{x,t+1}$. Similarly, the non-zero elements of the diagonal matrix $\mathbf{B}_t \in \mathbb{C}^{N \times N}$ are $b_{ii,t} = 1$ for all $i \in I_{x,t+1}$. The model parameters α_{ij} , λ_{ji} and \mathbf{Q} are assumed to be known.

B. Examples

Use of the transition probabilities λ_{ji} along with signal model (2) enables compact modeling of dynamically evolving sparsity patterns. The potential applications include MRI, spectrum sensing, direction of arrival estimation, frequency tracking etc.

As an initial example, consider a slowly varying sparsity pattern $I_{x,t}$ over T snapshots, following¹

$$\lambda_{ji} = \begin{cases} 0.90 & i = j \\ 0.05 & i = j \pm 1, \text{ if } j \notin \{1, N\} \\ 0.10 & i = j + 1, \text{ if } j = 1 \\ 0.10 & i = j - 1, \text{ if } j = N \\ 0 & |i - j| > 1 \end{cases}. \quad (4)$$

A realization of this process is illustrated in Figure 1. This choice is intended to model the strong temporal correlation of sparse signals exhibited in e.g. MRI.

Next, consider a simpler parameterization,

$$\lambda_{ji} = \begin{cases} 1 - \frac{N-1}{N}\nu & i = j \\ \frac{1}{N}\nu & i \neq j \end{cases}, \quad (5)$$

where $\nu \in [0, 1]$ is a mixture factor. This is intended to model more erratically evolving patterns in e.g. frequency-hopping radio frequency (RF) signals. Examples of resulting sparsity patterns are shown in Figure 2, where we consider $\nu = 0.01$ and 0.5 . It can be seen that the evolution of the sparsity pattern becomes more erratic as ν increases. In the above examples we have ensured that the sparsity level is constant, $K = 10$.

III. DYNAMIC ITERATIVE PURSUIT

We approach the dynamic estimation problem by first developing an iterative pursuit algorithm that can incorporate prior information in the form of a prediction of \mathbf{x}_t . A support set detection strategy is proposed using the signal to prediction error ratio. Next we develop a recursive algorithm based on the Kalman-filter framework. We propose predicting the sparse process as the superposition of all state transitions of the signal coefficients.

¹The two cases $j = 1$ or $j = N$ are necessary for the edge states.

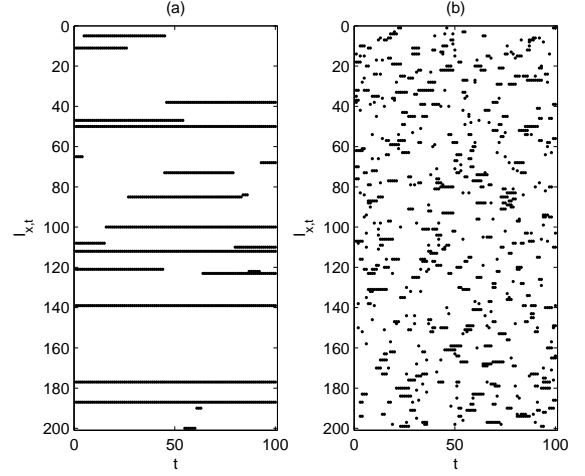


Fig. 2. Examples of evolving sparsity patterns with $(N, K, T) = (200, 10, 100)$ and transition probabilities (5) with (a) $\nu = 0.01$ and (b) $\nu = 0.5$.

A. Incorporation of prior information

Given the constraint on the support set, $|I_x| \leq K_{\max}$, the brute force least-squares solution would be to enumerate all combinations of possible support sets. For each set, $I \subset \{1, \dots, N\}$, the signal coefficients are reconstructed by a least-squares criterion and a measurement residual is computed, $\mathbf{r} = \mathbf{y} - \mathbf{H}_{[:,I]}\hat{\mathbf{x}}_{[I]}$. The reconstruction with minimum residual norm is then chosen as the solution. However, with at least one active coefficient there are $\sum_{K=1}^{K_{\max}} \binom{N}{K}$ possible support sets I to enumerate, which is clearly intractable.

Several iterative pursuit algorithms solve the estimation problem by a sequential detection of the support set and reconstruction of the corresponding signal coefficients. We will use OMP to illustrate the essential components of this sequential strategy.

OMP takes a support set I as its starting point. Reconstructed signal coefficients, \hat{x}_j , $j \in I$, are cancelled from the observation \mathbf{y} to form the residual $\mathbf{r} = \mathbf{y} - \mathbf{H}_{[:,I]}\hat{\mathbf{x}}_{[I]}$. Initially $I = \emptyset$. Under the hypothesis of a remaining active coefficient x_i , $i \notin I$, the residual signal model is

$$\mathbf{r} = \mathbf{h}_i x_i + \sum_{j \in I} \mathbf{h}_j \xi_j + \mathbf{n}, \quad (6)$$

where $\xi_j = x_j - \hat{x}_j$ are estimation errors. OMP detects the active coefficient by using a matched filter. The matched filter employs the strategy of estimating x_i using a least-squares criterion, $\tilde{x}_i = \mathbf{h}_i^\dagger \mathbf{r} = \mathbf{h}_i^* \mathbf{r}$. The index $i \notin I$ corresponding to maximum energy $|\tilde{x}_i|^2$ is added to I . Finally the coefficients corresponding to I are estimated jointly based on a least-squares criterion, solving

$$\hat{\mathbf{x}}_{[I]} = \arg \min_{\mathbf{x}_{[I]} \in \mathbb{C}^{|I|}} \|\mathbf{y} - \mathbf{H}_{[:,I]}\mathbf{x}_{[I]}\|_2^2 = \mathbf{H}_{[:,I]}^\dagger \mathbf{y}.$$

The residual \mathbf{r} is updated and the process is repeated until the residual norm no longer decreases or when $|I|$ reaches the limit K_{\max} . For sake of clarity OMP is summarized in Algorithm 1 where k denotes the iteration index.

Using a stochastic framework, we now extend the estimation strategy to a scenario in which a prediction $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{e}$ is given, where $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}^-)$ and error covariance matrix \mathbf{P}^- is known. Then the signal to prediction error ratio,

$$\rho_i \triangleq \frac{\mathbb{E}[|x_i|^2]}{\mathbb{E}[|e_i|^2]} \in [0, \infty), \quad (7)$$

quantifies the certainty that i belongs to the support set. We propose to use ρ_i for selecting indices to be added to I . The ratio ρ_i is

Algorithm 1 : Orthogonal Matching Pursuit (OMP)

```

1: Given:  $\mathbf{y}$  and  $\mathbf{H}$ 
2: Set  $k = 0$ ,  $\mathbf{r}_0 = \mathbf{y}$  and  $I = \emptyset$ 
3: repeat
4:    $k := k + 1$ 
5:    $i_k = \arg \max_{i \in I^c} |\mathbf{h}_i^* \mathbf{r}_{k-1}|$ 
6:    $I := I \cup i_k$ 
7:    $\hat{\mathbf{x}}_{[I]} = \mathbf{H}_{[:,I]}^\dagger \mathbf{y}$ ;  $\hat{\mathbf{x}}_{[I^c]} = \mathbf{0}$ 
8:    $\mathbf{r}_k = \mathbf{y} - \mathbf{H}_{[:,I]} \hat{\mathbf{x}}_{[I]}$ 
9: until ( $\|\mathbf{r}_k\|_2 \geq \|\mathbf{r}_{k-1}\|_2$ ) or ( $k > K_{\max}$ )
10: Output:  $\hat{\mathbf{x}}$  and  $I$ 

```

successively updated by conditioning the expectations on the residual, under the hypothesis with signal model (6). Then $E[|x_i|^2] = |\mu_{i|r}|^2 + \sigma_{i|r}^2$ where the conditional mean $\mu_{i|r}$ is given by the MMSE-estimator and $\sigma_{i|r}^2$ by its error variance. The prior of x_i is the prediction \hat{x}_i^- .

For tractability the estimation errors ξ_j are assumed to be Gaussian and their correlations negligible so that the MMSE-estimator gives

$$\begin{aligned} \mu_{i|r} &\simeq \hat{x}_i^- + \mathbf{g}_i^* (\mathbf{r} - \mathbf{h}_i \hat{x}_i^-) \\ \sigma_{i|r}^2 &\simeq (1 - \mathbf{g}_i^* \mathbf{h}_i) p_i^-, \end{aligned} \quad (8)$$

where p_i^- is the i th diagonal element of \mathbf{P}^- [17]. The gain (row) vector \mathbf{g}_i^* and covariance matrix \mathbf{D} are,

$$\begin{aligned} \mathbf{g}_i^* &= \left(\frac{1}{p_i^-} + \mathbf{h}_i^* \mathbf{D}^{-1} \mathbf{h}_i \right)^{-1} \mathbf{h}_i^* \mathbf{D}^{-1}, \\ \mathbf{D} &= \sum_{j \in I} \sigma_j^2 \mathbf{h}_j \mathbf{h}_j^* + \mathbf{R}, \end{aligned} \quad (9)$$

where σ_j^2 is the variance of ξ_j and \mathbf{R} is the covariance matrix of \mathbf{n} . As the support set I successively grows, the inverse \mathbf{D}^{-1} can be updated efficiently using the Sherman-Morrison formula, as shown below.

To sum up, the signal to prediction error ratio is given by

$$\rho_i = \frac{|\mu_{i|r}|^2 + \sigma_{i|r}^2}{p_i^-} \quad (10)$$

and approximated using (8). For the maximum ρ_i , $i \notin I$ is added to I . Finally, signal coefficients are jointly re-estimated, solving a weighted least-squares problem

$$\hat{\mathbf{x}}_{[I]} = \arg \min_{\mathbf{x}_{[I]} \in \mathbb{C}^{|I|}} \left\| \begin{bmatrix} \mathbf{y} \\ \hat{\mathbf{x}}_{[I]}^- \end{bmatrix} - \begin{bmatrix} \mathbf{H}_{[:,I]} \\ \mathbf{I}_{|I|} \end{bmatrix} \mathbf{x}_{[I]} \right\|_{\mathbf{R}^{-1} \oplus \mathbf{S}^{-1}}, \quad (11)$$

where $\mathbf{S} = \mathbf{P}_{[I,I]}^-$. This is the linear MMSE estimator provided I is the correct support set. The residual \mathbf{r} is updated and the process is repeated as above. The resulting algorithm is referred to as ‘Predictive OMP’ (PrOMP) and is summarized in Algorithm 2.

The function `mmse-rec` solves (11) and can be computed by a measurement update of form:

$$\hat{\mathbf{x}}_{[I]} = \hat{\mathbf{x}}_{[I]}^- + \mathbf{K} (\mathbf{y} - \mathbf{H}_{[:,I]} \hat{\mathbf{x}}_{[I]}^-),$$

where $\mathbf{K} = (\mathbf{S}^{-1} + \mathbf{H}_{[:,I]}^* \mathbf{R}^{-1} \mathbf{H}_{[:,I]})^{-1} \mathbf{H}_{[:,I]}^* \mathbf{R}^{-1}$. The function `update-cov` updates the inverse covariance matrix for the added reconstructed coefficient \hat{x}_i ,

$$\mathbf{D}^{-1} := \mathbf{D}^{-1} \left(\mathbf{I}_M - \frac{\mathbf{h}_i \mathbf{h}_i^* \mathbf{D}^{-1}}{\sigma_i^{-2} + \mathbf{h}_i^* \mathbf{D}^{-1} \mathbf{h}_i} \right),$$

where σ_i^2 is the corresponding diagonal element of the posterior error covariance matrix $(\mathbf{S}^{-1} + \mathbf{H}_{[:,I]}^* \mathbf{R}^{-1} \mathbf{H}_{[:,I]})^{-1}$ [17].

Algorithm 2 : Predictive Orthogonal Matching Pursuit (PrOMP)

```

1: Given:  $\mathbf{y}$ ,  $\mathbf{H}$ ,  $\mathbf{R}^{-1}$ ,  $\hat{\mathbf{x}}^-$  and  $\mathbf{P}^-$ 
2: Set  $k = 0$ ,  $\mathbf{r}_0 = \mathbf{y}$ ,  $I = \emptyset$  and  $\mathbf{D}^{-1} = \mathbf{R}^{-1}$ 
3: repeat
4:    $k := k + 1$ 
5:   Compute  $\rho_i$  using (10) and (8)
6:    $i_k = \arg \max_{i \in I^c} \rho_i$ 
7:    $I := I \cup i_k$ 
8:    $\hat{\mathbf{x}}_{[I]} = \text{mmse-rec}(\mathbf{y}, \mathbf{H}, \mathbf{R}^{-1}, \hat{\mathbf{x}}^-, \mathbf{P}^-, I)$ ;  $\hat{\mathbf{x}}_{[I^c]} = \mathbf{0}$ 
9:    $\mathbf{r}_k = \mathbf{y} - \mathbf{H}_{[:,I]} \hat{\mathbf{x}}_{[I]}$ 
10:   $\mathbf{D}^{-1} = \text{update-cov}(\mathbf{D}^{-1}, \mathbf{H}, \mathbf{R}^{-1}, \mathbf{P}^-, I, i_k)$ 
11: until ( $\|\mathbf{r}_k\|_2 \geq \|\mathbf{r}_{k-1}\|_2$ ) or ( $k > K_{\max}$ )
12: Output:  $\hat{\mathbf{x}}$  and  $I$ 

```

B. Robust support-set based strategy

The strategy described above performs a successive cancellation of reconstructed signal coefficients. The performance is therefore crucially dependent on detecting an active coefficient individually at each stage, which is a ‘hard’ decision. But the hypothesis of one remaining active coefficient at each stage induces a risk of irreversible detection errors. This increases with more erratically evolving sparsity patterns, since the process is harder to predict.

The signal to prediction error ratio ρ_i , however, provides a statistic that can be viewed as ‘soft information’. In order to increase robustness to detection errors we propose to use ρ_i for selecting the ℓ most likely remaining coefficients. Let us denote the set of ℓ most likely indices by L . It is joined with the existing set I to form a hypothesized support set $\tilde{I} = I \cup L$. This set is used to reconstruct $\hat{\mathbf{x}}_{[\tilde{I}]}$ and the coefficient \hat{x}_i , $i \in L$ with maximum magnitude is added to the support set I at each stage. Here $\ell = \max(0, K_{\max} - |I|)$, which prevents overfitting beyond the prior knowledge of the sparsity level.

Algorithm 3 describes this alternative detection strategy, based on a hypothesized support set. The concerned scheme is called ‘robust predictive OMP’ (rPrOMP).

Algorithm 3 : Robust predictive OMP (rPrOMP)

```

1: Given:  $\mathbf{y}$ ,  $\mathbf{H}$ ,  $\mathbf{R}^{-1}$ ,  $\hat{\mathbf{x}}^-$  and  $\mathbf{P}^-$ 
2: Set  $k = 0$ ,  $\mathbf{r}_0 = \mathbf{y}$ ,  $I = \emptyset$  and  $\mathbf{D}^{-1} = \mathbf{R}^{-1}$ 
3: repeat
4:    $k := k + 1$ 
5:   Compute  $\rho_i$  using (10) and (8)
6:    $\ell = \max(0, K_{\max} - |I|)$ 
7:    $L = \{\text{indices of } \ell \text{ largest } \rho_i \in I^c\}$ 
8:    $\tilde{I} = I \cup L$ 
9:    $\hat{\mathbf{x}}_{[\tilde{I}]} = \text{mmse-rec}(\mathbf{y}, \mathbf{H}, \mathbf{R}^{-1}, \hat{\mathbf{x}}^-, \mathbf{P}^-, \tilde{I})$ 
10:   $i_k = \arg \max_{i \in L} |\hat{x}_i|$ 
11:   $I := I \cup i_k$ 
12:   $\hat{\mathbf{x}}_{[I]} = \text{mmse-rec}(\mathbf{y}, \mathbf{H}, \mathbf{R}^{-1}, \hat{\mathbf{x}}^-, \mathbf{P}^-, I)$ ;  $\hat{\mathbf{x}}_{[I^c]} = \mathbf{0}$ 
13:   $\mathbf{r}_k = \mathbf{y} - \mathbf{H}_{[:,I]} \hat{\mathbf{x}}_{[I]}$ 
14:   $\mathbf{D}^{-1} = \text{update-cov}(\mathbf{D}^{-1}, \mathbf{H}, \mathbf{R}^{-1}, \mathbf{P}^-, I, i_k)$ 
15: until ( $\|\mathbf{r}_k\|_2 \geq \|\mathbf{r}_{k-1}\|_2$ ) or ( $k > K_{\max}$ )
16: Output:  $\hat{\mathbf{x}}$  and  $I$ 

```

C. Prediction of dynamic sparse signals

Suppose a snapshot \mathbf{y}_t has been observed and a prediction $\hat{\mathbf{x}}_t^-$ is given along with \mathbf{P}_t^- . Let $\hat{\mathbf{x}}_t$ denote the estimated sparse state vector after the application of a predictive greedy pursuit algorithm (either PrOMP or rPrOMP), and I its support set. Then the updated error covariance matrix \mathbf{P}_t is computed block-wise corresponding to the set I

and its complement I^c . First, $\mathbf{P}_{[I,I],t} = (\mathbf{S}_t^{-1} + \mathbf{H}_{[:,I]}^* \mathbf{R}_t^{-1} \mathbf{H}_{[:,I]})^{-1}$ is the posterior error covariance, where $\mathbf{S}_t = \mathbf{P}_{[I,I],t}^-$ [17]. Second, the uncertainty of the inactive coefficients is preserved by $\mathbf{P}_{[I^c,I^c],t} = \mathbf{P}_{[I^c,I^c],t}^-$. Finally, in line with the MMSE reconstruction (11), the cross-correlations are set as $\mathbf{P}_{[I,I^c],t} = \mathbf{0}$ and $\mathbf{P}_{[I^c,I],t} = \mathbf{0}$.

We propose predicting \mathbf{x}_{t+1} from $\hat{\mathbf{x}}_t$ as a superposition of all possible transitions,

$$\hat{\mathbf{x}}_{t+1}^- = \sum_{j=1}^N \lambda_{ji} \alpha_{ij} \hat{\mathbf{x}}_{j,t}, \quad (12)$$

or written compactly, $\hat{\mathbf{x}}_{t+1}^- = \mathbf{F} \hat{\mathbf{x}}_t$, where $f_{ij} = \lambda_{ji} \alpha_{ij}$. The prediction error covariance matrix is then approximated by the equation, $\mathbf{P}_{t+1}^- = \mathbf{F} \mathbf{P}_t \mathbf{F}^* + \mathbf{Q}$.

Putting these blocks together we develop a Kalman-filter based algorithm for recovery of sparse processes in Algorithm 4, which we call dynamic iterative pursuit (DIP). In DIP we use predictive OMP (PrOMP). If robust predictive OMP (rPrOMP) is used instead, the algorithm can be referred to as ‘rDIP’.

Algorithm 4 : Dynamic Iterative Pursuit (DIP)

```

1: Initialization  $\hat{\mathbf{x}}_0^-$  and  $\mathbf{P}_0^-$ 
2: for  $t = 0, \dots$  do
3:   %Measurement update
4:    $[\mathbf{x}_t, I] = \text{PrOMP}(\mathbf{y}_t, \mathbf{H}, \mathbf{R}_t^{-1}, \hat{\mathbf{x}}_t^-, \mathbf{P}_t^-)$ 
5:    $\mathbf{S}_t = \mathbf{P}_{[I,I],t}^-$ 
6:    $\mathbf{P}_{[I,I],t} = (\mathbf{S}_t^{-1} + \mathbf{H}_{[:,I]}^* \mathbf{R}_t^{-1} \mathbf{H}_{[:,I]})^{-1}$ 
7:    $\mathbf{P}_{[I^c,I^c],t} = \mathbf{P}_{[I^c,I^c],t}^-$ ;  $\mathbf{P}_{[I,I^c],t} = \mathbf{0}$ ;  $\mathbf{P}_{[I^c,I],t} = \mathbf{0}$ 
8:   %Prediction
9:    $\hat{\mathbf{x}}_{t+1}^- = \mathbf{F} \hat{\mathbf{x}}_t$ 
10:   $\mathbf{P}_{t+1}^- = \mathbf{F} \mathbf{P}_t \mathbf{F}^* + \mathbf{Q}$ 
11: end for
```

IV. EXPERIMENTS AND RESULTS

In this section we evaluate DIP with respect to static OMP, SP and convex relaxation based basis pursuit denoising (BPDN) [1] algorithms. We also show the performance of a ‘genie-aided’ Kalman filter (KF) which provides a bound for MMSE-based reconstruction of linear processes. The genie-aided approach is given the sparsity pattern a priori, but does not know the active signal coefficients. Finally, the robustness properties of rDIP are compared with DIP for erratically evolving sparsity patterns. The results are shown using Monte Carlo simulations, averaged over 100 runs.

A. Signal generation and performance measure

Using a typical setup we consider a sparse process with the parameters $N = 200$, $K = 10$ and number of snapshots $T = 100$, with oscillating coefficients according to an AR-model as in (2) with $\alpha_{ij} = \alpha \equiv -0.8$, and $\mathbf{Q} = \sigma_w^2 \mathbf{I}_N$. The sparsity pattern transitions, $I_{x,t} \rightarrow I_{x,t+1}$, are determined by transition probabilities λ_{ji} which are set in the experiments.

The transition of each active state $j \in I_{x,t}$ is generated by a first-order Markov chain with λ_{ji} . If two states in $I_{x,t}$ happen to transition into one, a new state is randomly assigned to $I_{x,t+1}$, to ensure that the sparsity level is constant in the experiment.

The entries of the sensing matrix \mathbf{H} are set by random drawing from a Gaussian distribution $\mathcal{N}(0,1)$ followed by unit-norm column scaling. The measurement noise covariance matrix has form $\mathbf{R}_t = \sigma_n^2 \mathbf{I}_M$. Process and measurement noise are generated as $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $\mathbf{n}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$, respectively.

In the experiments, two signal parameters are varied; (a) the signal-to-measurement noise ratio,

$$\text{SMNR} \triangleq \frac{\mathbb{E} [\sum_t \|\mathbf{x}_t\|_2^2]}{\mathbb{E} [\sum_t \|\mathbf{n}_t\|_2^2]}, \quad (13)$$

while fixing $\mathbb{E} [\|\mathbf{x}_t\|_2^2] \equiv 1$ so that $\sigma_n^2 = \frac{1}{M \times \text{SMNR}}$, and (b) the fraction of measurements $\kappa \triangleq M/N$.

For a performance measure we use the signal-to-reconstruction error ratio, defined as

$$\text{SRER} \triangleq \frac{\mathbb{E} [\sum_t \|\mathbf{x}_t\|_2^2]}{\mathbb{E} [\sum_t \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2^2]}, \quad (14)$$

which is the inverse of the normalized MSE. Note that $\text{SRER} = 0$ dB, i.e. no reconstruction gain, is equivalent to using $\hat{\mathbf{x}}_t = \mathbf{0}$.

B. Algorithm initialization

For the predictive algorithms—DIP, rDIP and genie-aided KF—we use the mean and variance of an autoregressive process as initial values, $\hat{\mathbf{x}}_0^- = \mathbf{0}$ and $\mathbf{P}_0^- = \sigma_x^2 \mathbf{I}_N$ where $\sigma_x^2 = \frac{\sigma_w^2}{1-\alpha^2}$. In these algorithms we set $K_{\max} = K$ for consistent comparisons, although strict equality is not a requirement.

Here we mention that BPDN [1] solves

$$\hat{\mathbf{x}}_t = \arg \min_{\mathbf{x}_t \in \mathbb{R}^N} \|\mathbf{x}_t\|_1 \text{ subject to } \|\mathbf{y}_t - \mathbf{H} \mathbf{x}_t\|_2 \leq \varepsilon,$$

where the slack parameter ε is determined by the measurement noise power, as

$$\varepsilon = \sqrt{\sigma_n^2 (M + (2\sqrt{2M}))},$$

following [18], [16]. Note that BPDN does not provide a K -element solution. It is also unable to use prediction. The code for BPDN is taken from the l_1 -magic toolbox.

C. Results

For all experiments we ran 100 Monte Carlo simulations, where a new realization of $\{\mathbf{x}_t, \mathbf{y}_t\}_{t=1}^T$ and \mathbf{H} was generated for each run.

In the first experiment we consider a slowly varying sparsity pattern $I_{x,t}$ following the transition probabilities λ_{ji} in (4). Figure 3 shows how the algorithms perform with varying measurement noise power at a fixed fraction of measurements $\kappa = 0.25$. DIP overtakes static BPDN at lower SMNR levels, while exhibiting a similar graceful degradation. The static OMP and SP do not take into account the measurement noise and hence continue to degrade. For instance, DIP reaches the cut-off point of 0 dB reconstruction gain at an SMNR level that is approximately 5 dB lower than the static OMP. Figure 4 shows how the improvements persist for varying κ at a fixed SMNR = 10 dB. In this scenario rDIP exhibits similar performance as DIP. Taking static OMP as the baseline algorithm, the improvement of predictive iterative pursuit is illustrated in Figure 5. The minimum SRER advantage is about 2 dB, and increases substantially with rising SMNR.

Next, we consider an unknown but static sparsity pattern $I_{x,t} \equiv I_x$, generated by degenerate transition probabilities $\lambda_{ji} = \delta(i - j)$, and compare DIP with a ‘genie-aided’ KF. The latter filters the coefficients of a known support set $I_{x,t}$, and therefore provides an upper bound on the performance of sequential estimation. The bound is not necessarily tight since only $|I_{x,t}| \leq K$ is given in the problem. As SMNR increases, DIP rapidly approaches the bound while OMP saturates for $\kappa = 0.25$, illustrated in Figure 6. At SMNR = 20 dB, OMP and DIP are about 10 and 2 dB from the upper limit, respectively. Again, rDIP performs similarly to DIP.

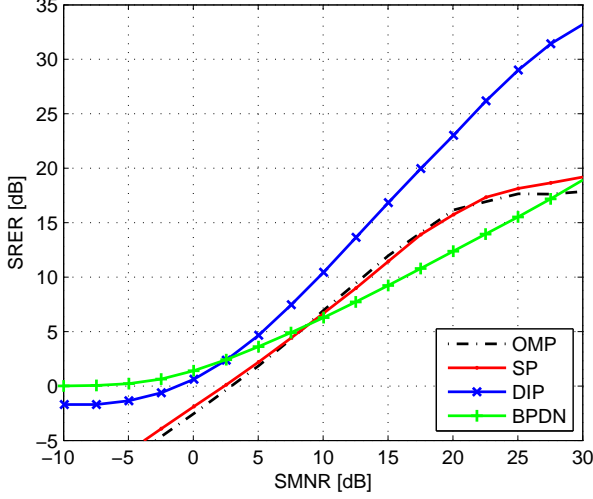


Fig. 3. Comparison of different methods where we show SRER versus SMNR at $\kappa = 0.25$ and set transition probabilities according to (4).

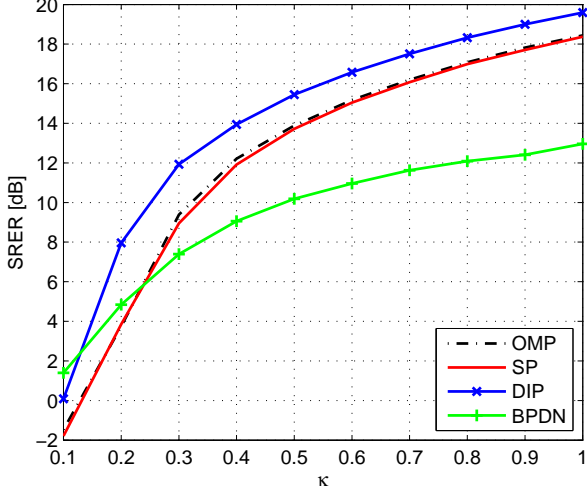


Fig. 4. SRER versus $\kappa = M/N$ at SMNR = 10 dB. Transition probabilities according to (4).

Finally, we consider an erratically evolving sparsity pattern $I_{x,t}$, with transition probabilities λ_{ji} set according to (5), in order to compare the robustness of rDIP with DIP. Figure 7 shows how performance is affected as the mixture factor ν increases. DIP converges to OMP from above; rDIP provides near equivalent performance to DIP at first but shows a more graceful degradation. At the extreme, when all transitions are equiprobable, rDIP is still capable of yielding above +2.5 dB gain over OMP. This validates the robustness considerations behind its design.

Reproducible results: MATLAB code for the algorithms can be downloaded from <http://sites.google.com/site/saikatchatt/softwares>. The codes produce the results in Figures 3 and 7.

V. CONCLUSIONS

We have developed a new iterative pursuit algorithm that uses sequential predictions for dynamic compressive sensing, which we call dynamic iterative pursuit. It incorporates prior statistical information using linear MMSE reconstruction and the signal to prediction error as a statistic. The algorithm was experimentally tested on a sparse

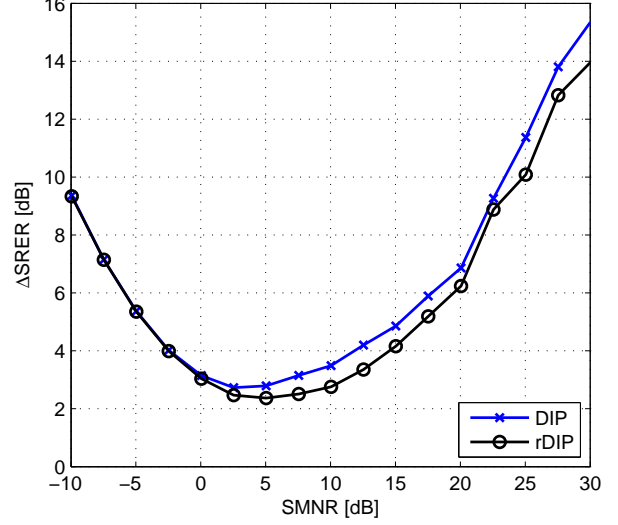


Fig. 5. SRER improvement over OMP versus SMNR at $\kappa = 0.25$. Transition probabilities according to (4).

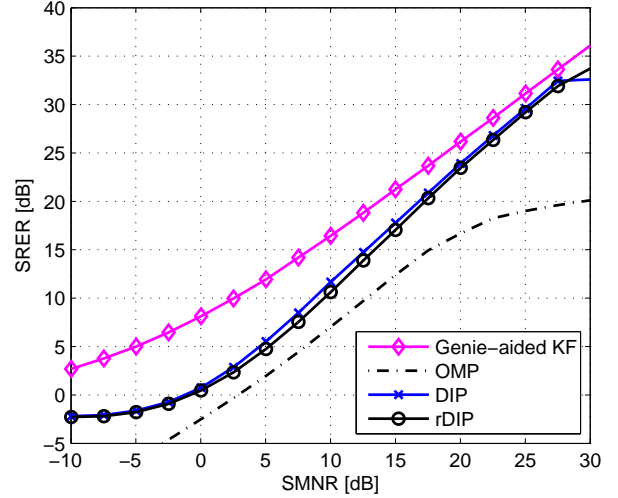


Fig. 6. SRER versus SMNR at $\kappa = 0.25$. Degenerate transition probabilities (fixed sparsity pattern).

signal with oscillating coefficients and evolving sparsity pattern. The results show that the algorithms exhibit graceful degradation at low SMNR regions while capable of yielding substantial performance gains as the SMNR level increases.

VI. ACKNOWLEDGEMENT

The authors would like to thank E. Candes and J. Romberg making the l_1 -magic toolbox available online.

REFERENCES

- [1] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, Mar. 2008.
- [2] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 53, pp. 4655–4666, Dec. 2007.
- [3] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Information Theory*, vol. 55, pp. 2230–2249, May 2009.

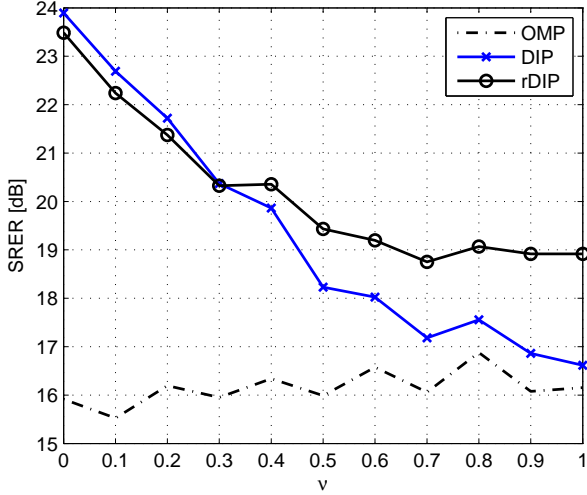


Fig. 7. SRER versus mixture factor ν at SMNR = 20 dB and $\kappa = 0.25$. Transition probabilities according to (5).

- [4] D. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," tech. rep., 2006.
- [5] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
- [6] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Computational Mathematics*, vol. 9, pp. 317–334, 2009.
- [7] S. Chatterjee, D. Sundman, and M. Skoglund, "Look ahead orthogonal matching pursuit," in *IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP)*, May 2011.
- [8] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [9] N. Vaswani, "LS-CS-residual (LS-CS): Compressive sensing on least squares residual," *IEEE Trans. Signal Processing*, vol. 58, pp. 4108–4120, Aug. 2010.
- [10] D. Sundman, S. Chatterjee, and M. Skoglund, "On the use of compressive sampling for wide-band spectrum sensing," in *IEEE Intl. Symp. on Signal Processing and Information Technology (ISSPIT)*, pp. 354–359, Dec. 2010.
- [11] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.
- [12] N. Vaswani, "Kalman filtered compressed sensing," in *15th IEEE Int. Conf. on Image Processing (ICIP) 2008.*, pp. 893–896, Oct. 2008.
- [13] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Trans. Signal Processing*, vol. 58, pp. 2405–2409, Apr. 2010.
- [14] Z. Zhang and B. Rao, "Sparse signal recovery in the presence of correlated multiple measurement vectors," in *IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP), 2010*, pp. 3986–3989, Mar. 2010.
- [15] W. Dai, D. Sejdinovic, and O. Milenkovic, "Gaussian dynamic compressive sensing," in *Int. Conf. Sampling Theory and Applications (SampTA)*, May 2011.
- [16] S. Chatterjee, D. Sundman, and M. Skoglund, "Robust matching pursuit for recovery of Gaussian sparse signal," in *DSP Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*, pp. 420–424, Jan. 2011.
- [17] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.
- [18] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.